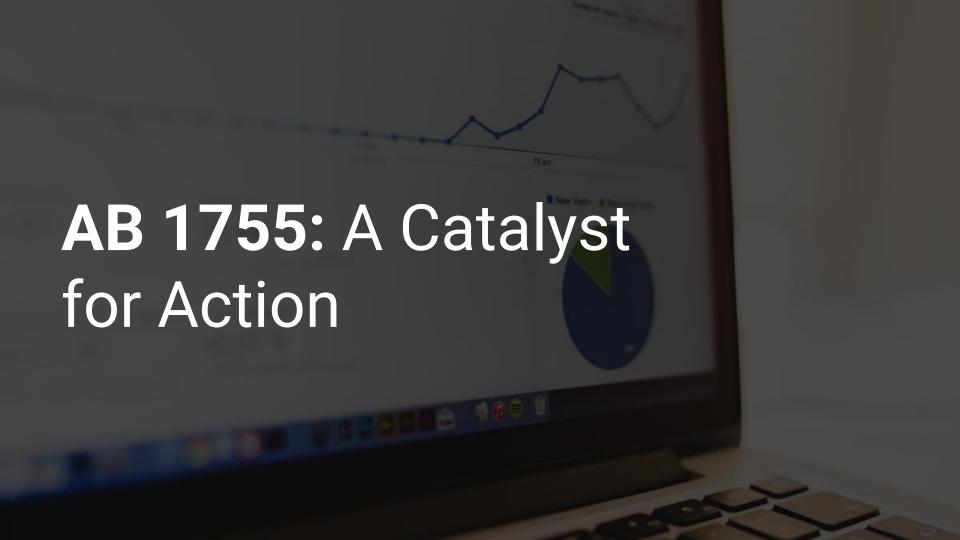
The Transition to Open Data

Where to go from here...

Tony Hale San Francisco Estuary Institute Feb 28, 2017

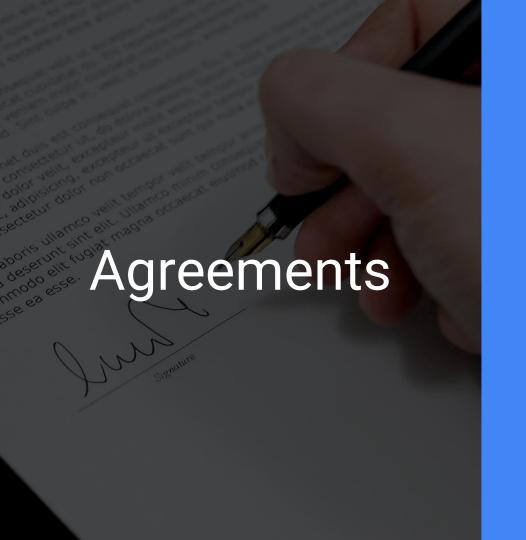


Opportunities and Hazards

Open data is great. Open data is necessary. But it is not without its perils.

"Data is all very well, but the ability to extract meaningful information from it requires considerable skill. Without it, there is a real danger incorrect conclusions may be drawn."

-- Prof. David Hand, Imperial College



Memoranda of Agreement

Sharing among agencies

Proposed use cases reveal need for interagency sharing

"MOAs can facilitate the sharing of data."

-- Zach Townsend, CDO

What is a Digital Object Identifier?

pe: Journal Article

- Invented in 2000.
- Definition: "A digital object identifier (DOI) is a unique alphanumeric string assigned to a publication or dataset."
 - Examples:

http://ascelibrary.org/doi/abs/10.1061/(ASCE)EE.1943-

7870.0000921

http://www.doi.org/demos.html

Why is this an urgent matter?

Open Data, as practiced in California, is about transparency and data mobility

Mobile data is unmoored from its origins

The ascendancy of "open data" could lead to the rise of bad data:

Data without attribution

Data without a chain of custody

Data that might be inadvertently duplicated

Data that might be misused/misunderstood as the data producers are further divorced from data consumers

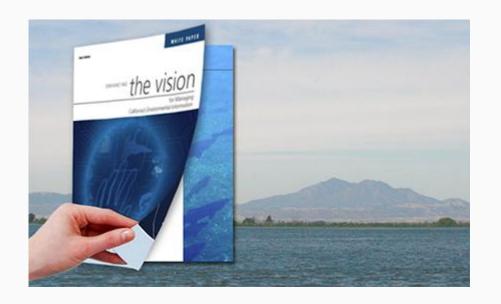
Connections to the Vision Document

Provenance / Data Attribution

Metadata

Data governance

Discoverability



The Way Forward for California

Scenario to protect data integrity and associated investments:

Every repository has a place for a DOI

Documentation about DOIs is available (education)

DOI system is universally recognized

Every dataset that is downloaded has an







Requisite Resources

- Modifications to repositories: \$? Simple operation
- Modifications to distribution applications: \$? Slightly more complicated
- Development of documentation: Fairly cheap
- Establish accounts with a DOI Authority: Modest cost
- Generate DOIs for current and future datasets with authorized provider: Incurs modest cost
 - Can be performed individually
 - Can be performed in bulk (via scripts)

Questions?

Tony Hale

tonyh@sfei.org